Final Project

Andriy Kashyrskyy, Marta Zaremba

CS112 Fall 2020

Decision Memo

To: President and CEO of *La Fundación Internacional para los Sistemas Electorales - República Argentina (The International Foundation for Electoral Systems - Argentina)*

From: External Advisors

Question: Is e-voting likely to improve the voting experience among citizens in Argentina during the upcoming Senate elections in October 2021?

Proposed Actions: (1) Analysis of the effectiveness of e-voting on voters' satisfaction (confidentiality, duration, ease of voting, etc.) via statistical matching with past voting data, (2) implementation of encouragement design for the upcoming Senate elections in Argentina.

Summary

With the goal of increasing the quality and transparency of the governmental elections, states in South America are currently analyzing the effectiveness of changing the voting technology. The national study in Salta, Argentina, "Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions" by Alvares et al. (2013), examined the influence of voting type on voters' experience. By conducting an observational study and comparing the voting experience of e-voters with traditional voters, researchers found that e-voters are more likely to report this type of voting as easier, have a higher certainty of votes being counted, and support the replacement of traditional voting by e-voting. E-voters, however, tend to be more concerned about the ballot's secrecy than traditional voters.

In this paper, we replicate the original findings of the estimated treatment effect of e-voting on voter's experience compared to traditional voting. The results of the replication are consistent with the previous findings. Additionally, we aim to improve the original finding's accuracy by changing the method of estimation of the causal treatment effect. We find that the proposed method does not improve the results from an original one, altering the treatment effect estimation. We build on the study's finding on the effectiveness of e-voting and extend it with the methodology of encouragement design, potentially to be implemented in the national Argentinian context for the upcoming Senate Elections in 2021.

Study design (*Data, methodology, and results)*

The original study uses data of 1502 voters from Salta, Argentina, collected on the election day in 2011. 887 people used e-voting, and the rest (615) voted traditionally. At the poll station, the voters responded to a survey, evaluating their voting experience and providing demographic information on: age, education, whether they work in white collar, gender, technology count, and political information. There were 9 *binary outcome variables* that identified the evaluation of the voting's effectiveness, and these can be found in Table 1 in Appendix A.

A single individual could not participate in both types of voting at once, and thus it was impossible to compute the true individual causal effect[1]. Besides, the treatment (e-voting) in the study was not assigned randomly[2]; the government chose polling places with "greater populations of technologically sophisticated, better-educated and higher-income voters."[3]

The researchers addressed solving the non-randomization problem by conducting statistical matching and pairing individuals of the treatment group (e-voters) with the similar ones in the control group (traditional voters) based on a set of observed characteristics. The study used propensity score matching, pairing individuals based on a score (probability of being an e-voter) rather than each of the covariates (characteristics). The researchers also specified that the scores between the voters could not differ more than 5 percent[4]. Matching is done to improve the balance of units' characteristics in the control and treatment group, making the groups similar to each other and ensuring that the only difference between two groups is the treatment type, which resembles a Randomized Control Trial (RCT). After matching, the causal treatment effect was calculated by defining the difference in the proportion of voting experience scores between the treatment and the control group (Appendix A: Table 3).

---

[1] Here we refer to **the Fundamental Problem of Causal Inference** that states that it is **not** possible to observe the causal effect on a single unit; we can observe only one of Y(1) or Y(0), Y being the outcome, with 0 (no treatment), and with 1 (treatment).
[2] Alvarez, *Voting made safe and easy: the impact of e-voting on citizen perceptions*, 118.
[3] Alvarez, 121.
[4] By setting the **caliper** argument to 0.05 when running the matchit() function in R software (code in Appendix E).

The results showed that the experience of e-voters improved significantly among most outcome variables, while the difference in speed of voting process evaluation was not significant. Besides, e-voters were less likely to be confident in ballot secrecy than traditional voters.

Replication of the study

To assess the reproducibility of the propensity score matching results, we replicate Table 2 to compare the balance before and after matching. We use the same method as in the original paper, using the logistic regression model defined in Appendix B.

The propensity score matching method is based on the multivariable adjustment, giving single scores for each observation (person) in the treatment and control group, and later matching the scores of units from the treatment group with the similar score of units from the control group. The score corresponds to the probability of being assigned to the treatment, in our case - to e-voting, thus it ranges between 0 and 1.

Upon replication of Table 2 (Balance Statistics), using the propensity score matching, we get the matched dataset with the same amount of matched observations (n = 1,164) as in the study; however, the post-matching replicated values (“educ”, “white.collar,” “not.full.time”, “male,” “tech,” “pol.info”) differ from the original paper (Appendix C). The slight difference in replication results (those of propensity score matching in replicated Table 2) might be the effect of randomness employed in the algorithm of propensity score matching, and due to using data and code from 2013, which was computationally run in the older version of R Software. Although slight differences in the replication of the propensity score remain, these are minimal and have almost no effect on estimating the causal treatment effect on e-voters.

Extension of the analysis: Methods

To compare the balance achieved through propensity score matching with another matching type, we conduct genetic matching. Propensity score matching was previously reported to worsen the covariates' balance in studies since creating good propensity score matching requires iteratively checking the improvement of the balance (Diamond & Sekhon, 2013). Diamond and

Sekhon (2013) observe many researchers not applying iterative propensity score matching, which can also be seen in *"Voting made safe and easy: the impact of e-voting on citizen perceptions*," with the propensity score matching ran only once.

In comparison, genetic matching automates the iterative process of repeatedly checking the balance, and it assigns different weights of importance for the covariates and outputs the ones that give the best balance. Afterward, the weights are used for matching the covariates, aiming to improve the variable with the lowest statistical p-value[5] in the balance between the treatment and control groups. Additionally, we make balance plots of covariates of treatment and control groups before and after propensity score matching and genetic matching (Appendices G, H).

Results

After conducting genetic matching, we notice that the balance for all the covariates besides "political information" worsened, as shown on the balance plots (Appendix G). We see that the lowest p-value (here this value shows how close the treatment and control groups are) under propensity score matching is 0.31 for "tech" (Table 2, Appendix C), under the genetic matching, it is 0.0001 for "educ" (Table 2, Appendix D). Such results inform us about possible pitfalls of the genetic algorithm: genetic matching aims to improve the balance of the least-balanced covariates at the cost of worsening the balance of the covariates that might have been balanced well before matching. Besides, to find the best possible balance, one needs to adjust the algorithm's arguments iteratively.

We also extend Table 3 to see how the results obtained through different matching types influence the estimation of the causal effect of e-voting. Comparing Table 3 under propensity score matching and genetic matching, one can see that the causal effect changed slightly. However, the conclusion of e-voting improving voters' experience among all the outcomes besides the confidence of ballot secrecy remains the same as in the original method.

[5] Here we note aiming to improve the **p-value**, with a bigger p-value in this context meaning rejection of the null hypothesis (that *there is no balance between treatment (Y1) and control (Y0) group*).

By conducting the extension of estimating the treatment effect after genetic matching, we notice that the type of matching and the balance achieved influence the study's finding. Therefore, we suggest the researcher use multiple types of matching to achieve the best balance before making conclusions after testing a single type. This is particularly crucial for the outcomes that have a small difference between the treatment and control group. Additionally, it would be useful to address the limitations and assumptions associated with the used methods and how these might influence the interpretation of results; we proceed to do so in the next section.

Suggestion for the future study

The original study data was collected in 2011 from Salta, Argentina, chosen as the study's location due to the high number of educated people with higher income and use of technology. In addition to the COVID-19 pandemic, since that time, Argentina has faced political and economic crises (Flannery, 2020) that might have influenced people's perception of their voting experience since the survey was conducted in 2011. Therefore, the estimation of the effect of e-voting on voting experience is expected to differ this year. Besides, as reported by the researchers of the original paper, the results of the study might not be generalizable to the whole Argentinian population since it was affected by significant selection bias in defining the locations that would be assigned to e-voting.[6] Therefore, we suggest conducting a new study that would allow collecting more recent data. The case of upcoming Senate elections in Argentina in October 2021 can serve as an opportunity for such a study.[7]

Since all citizens are free to choose the type of voting and cannot be forcibly assigned or denied to vote in a certain way, conducting a randomized controlled trial (RCT) would be inappropriate. Additionally, since most people in Argentina are familiar with traditional voting rather than e-voting, it is likely to attract fewer voters at the beginning. Therefore we suggest

[6] Alvarez, 121.

[7] **#gapanalysis:** we identified the gap in the available estimation of the treatment effect of e-voting and justified it by referring to political and economic issues that Argentina faced since the last results were estimated. To address the gap, we propose conducting an encouragement design study and explain why it would be appropriate.

conducting an *encouragement design study* known for being accessible and cost-efficient to implement. Under this design, rather than being randomly assigned to e-voting or traditional voting, people are randomly assigned to *encouragement*, which can take the form of incentive or information. This way, the treatment group is the one that receives the encouragement, and the control group does not. In the case of voting in elections, we suggest sending out booklets with information on e-voting to increase the individual likelihood of those in the encouragement group to e-vote. Since booklets can be easily distributed by mail, we suggest sending them to the randomly selected (voting-eligible) population in the locations with access to e-voting as of October 2021, as its equipment is not available in all regions yet. However, we have to note that this limits the generalizability of the causal treatment effect since the original study reported a selection bias of e-voting machines installment - populations in regions with e-voting machines have specific characteristics (e.g., high income, education, familiarity with technology), which might differ for the rest of Argentinian population.

An effective encouragement should increase the individual likelihood of e-voting; however, we expect not all people receiving the booklet to abide by the encouragement, and the ones who do not receive encouragement can still e-vote. When deciding on the encouragement, it is necessary to ensure that it affects everyone in the *same way* (e.g., everyone has to equally understand the language used for it), *does not* make people less likely to e-vote (e.g., making the process sound complicated may discourage people to e-vote), and *does not* directly affect the outcome (e.g., should not influence the assessment of voting experience).

On the day of the election (after encouragement design was implemented), it is necessary to distribute surveys among the voters to assess their voting experience based on the same questions and outcomes as described in the original study, which will enable comparison of the causal treatment effect on voting process perception of e-voters and traditional voters.

The limitation of estimating the causal effect of e-voting through such methods is that it calculates the local average treatment effect[8], meaning the treatment effect of people who abided by their assignment (voted electronically when receiving the booklet, and voted traditionally when not receiving it). This estimation can be problematic when there is a big proportion of people who did not comply with their assignment. Besides, the encouragement has to lead to a higher probability of being in treatment (e-voting) for the method to work. Lastly, e-voting has to be unfamiliar to most people in the region for encouragement to work, and it is expected in the Argentinian context since the majority had not voted electronically.

Discussion

With the need to increase the transparency of the upcoming Senate elections in Argentina, implementing e-voting can improve voters' experience and lead to higher satisfaction as seen from previously computed causal treatment effects (via propensity score and genetic matching).

To define whether e-voting causes the overall better experience of voting for Argentinians with a number of outcomes related to its ease, speed, confidentiality, cleanness, and more, we replicate the original study by running the propensity score matching, enabling to compare outcomes between balanced treatment and control groups. While we note, by looking at the balance plots and Table 2, the groups are becoming more balanced after matching with propensity scores (Appendix C & Appendix G). However, due to its frequent misapplication pointed by statisticians, we extend the prior analysis with genetic matching. Genetic matching, however, leads to a worse balance of covariates, as can be seen from balance plots and Table 2 (Appendix D & Appendix H). We have to note that we are only looking at the observed covariates, and unobserved ones are not included; therefore, the sensitivity analysis performed in the original paper examining the robustness of the model is suggested to be used after genetic matching.

---

[8] Here we refer to **LATE** (local average treatment effect), find the computational explanation in Appendix F.

With the outcomes of propensity score matching and genetic matching analyses (balanced treatment and control groups), we thus conclude that e-voting has a significant effect on the voting experience of Argentinians, leading to higher voting satisfaction, which includes higher voting speed, ease of procedure, the confidence of the vote being counted, confidence on ballot secrecy, cleanness of elections. However, e-voters are more likely to be uncertain about the ballot's secrecy, suggesting that the government should focus on educating the population about the details of the process of e-voting. These results are based on the data from 2011; therefore, before *The International Foundation for Electoral Systems - Argentina* decides to support the implementation of policies to establish e-voting on a larger scale during the Senate elections in 2021, we suggest conducting an encouragement design and estimating the causal treatment effect using more recent data.

Appendix

Appendix A: Outcome variables in the original study (Table 1)

Name of the variable	**Responses**
Evaluation of voting experience	1 if voting experience was 'very good'; and 0 if 'good', 'bad' or 'very bad'
Qualification of poll voters	1 if 'very qualified' or 'somewhat qualified'; and 0 if 'little qualified' or 'not at all qualified'
Substitution of traditional voting by e-voting	1 if 'agree a lot' or 'agree'; and 0 if 'disagree' or 'disagree a lot'
Speed of voting process	1 if 'very quick' or 'quick'; and 0 if 'slow' or 'very slow'
Ease of voting procedure	1 if voting was 'very easy'; and 0 if 'easy', 'difficult' or 'very difficult'
Preferred method for selecting candidates from different political parties	1 if 'electronically', and 0 if 'by hand'
Confidence on vote being counted	1 if 'very sure' or 'sure'; and 0 if 'unsure' or 'very unsure'
Confidence on ballot secrecy	1 if 'very confident' or 'confident'; and 0 if 'not

	confident' or 'not at all confident'
Cleanness of elections in Salta	1 if 'very clean' or 'somewhat clean'; and 0 if 'a little clean' or 'not at all clean'

Appendix B: Original study (logistic regression model, Table 2, Table 3)

Outcome variables:

- "capable.auth", eval.voting", "easy.voting", "sure.counted", "conf.secret", "how.clean", "speed", "agree.evoting", "eselect.cand"

Covariates:

- "age.group", "educ", "white.collar", "not.full.time", "male", "tech", "pol.info"

Logistic regression model

$$EV = age.group + I(age.group^2) + I(age.group^3) + age.group : educ + age.group : tech + educ + I(educ^2) + tech + I(tech^2) + pol.info + educ : pol.info + age.group : pol.info + tech : pol.info + white.collar + not.full.time + male$$

Table 2 - Balance Statistics:

TABLE 2 *Balance Statistics*

	Before matching (N = 1,475)				After matching (N = 1,164)				
	EV	TV	Diff.	p-value*	EV	TV	Diff.	p-value*	% Imp.
Age group (1–5)	2.5	2.4	0.0	0.55	2.5	2.5	0.0	1.00	100%
Education (1–8)	4.8	4.1	0.6	0.00	4.2	4.2	0.0	0.72	98%
White collar (%)	30.3	27.6	2.7	0.29	29.2	28.4	0.9	0.80	68%
Not full time worker (%)	27.7	33.5	−5.8	0.02	30.8	32.0	−1.2	0.80	79%
Male (%)	49.7	49.1	0.6	0.87	49.0	49.0	0.0	1.00	100%
Technology count (1–6)	4.2	3.9	0.3	0.00	4.0	3.9	0.1	0.59	76%
Political information (1–4)	1.5	1.3	0.2	0.00	1.4	1.3	0.0	0.55	77%

Note: For more details about balance improvement, see Online Appendix 3.
*p-values of Kolmogorov-Smirnov tests (ordinal variables), and difference in proportions tests (binary variables).

Table 3 - Causal Effect of e-voting:

TABLE 3 *Causal Effect of e-voting*

	Before matching (N = 1,475)					After matching (N = 1,164)				
	N	E-Voting (%)	Traditional Voting (%)	Diff.	p-value*	N	E-Voting (%)	Traditional Voting (%)	Diff.	p-value*
Select candidates electronically	1,388	83.8	53.4	30.4	0.000	1,101	82.7	54.1	28.6	0.000
Evaluation of voting experience	1,460	46.3	21.3	25.0	0.000	1,151	45.6	20.9	24.7	0.000
Ease of voting procedure	1,469	33.6	11.5	22.1	0.000	1,159	32.5	11.9	20.6	0.000
Agree substitute TV by EV	1,409	84.1	62.4	21.7	0.000	1,114	82.4	63.3	19.1	0.000
Elections in Salta are clean	1,284	58.0	41.0	17.0	0.000	1,022	57.6	41.5	16.0	0.000
Sure vote was counted	1,418	86.3	77.0	9.3	0.000	1,117	85.7	77.0	8.8	0.000
Qualification of poll workers	1,416	85.1	76.2	8.9	0.000	1,123	84.5	76.0	8.5	0.000
Speed of voting process	1,443	84.1	80.8	3.2	0.130	1,137	83.2	80.7	2.5	0.306
Confident ballot secret	1,431	77.1	84.5	−7.4	0.001	1,133	76.9	84.3	−7.4	0.002

Note: sample sizes given in the first column (before matching) differ from those given in Table 1 because Table 3 omits respondents with missing values in covariates.
*Test of difference in proportions.

Appendix C: Study replication (Propensity Score Matching)

Table 2 - Balance Statistics:

	ev	tv	diff	pvalue	ev	tv	diff	pvalue
age.group	2.5	2.4	0.032	0.570	2.5	2.5	-0.0017	1.00
educ	4.8	4.1	0.629	0.000	4.2	4.2	0.0292	0.49
white.collar	30.3	27.6	2.668	0.293	29.4	27.3	2.0619	0.47
not.full.time	27.7	33.5	-5.784	0.020	30.6	32.1	-1.5464	0.61
male	49.7	49.1	0.557	0.875	49.0	49.8	-0.8591	0.81
tech	4.2	3.9	0.274	0.000	4.0	3.9	0.0808	0.31
pol.info	1.5	1.3	0.164	0.002	1.4	1.3	0.0361	0.63

Table 3 - Causal Effect of e-voting:

	N	prop.ev	prop.tv	diff	pvalue	N	prop.ev	prop.tv	diff	pvalue
eselect.cand	1388	83.84	53.42	30.428	1.237e-34	1102	81.69	53.39	28.293	1.918e-23
eval.voting	1460	46.33	21.30	25.035	1.833e-22	1151	45.23	21.43	23.805	1.866e-17
easy.voting	1469	33.64	11.53	22.111	5.420e-22	1159	31.95	11.38	20.572	3.449e-17
agree.evoting	1409	84.14	62.44	21.705	2.864e-20	1113	82.76	62.59	20.175	7.122e-14
how.clean	1284	57.97	40.99	16.980	2.561e-09	1021	57.00	41.65	15.349	1.288e-06
capable.auth	1416	85.14	76.25	8.889	2.954e-05	1123	84.67	76.52	8.150	7.512e-04
sure.counted	1418	86.35	77.02	9.332	7.444e-06	1117	85.56	77.52	8.044	7.025e-04
speed	1443	84.06	80.85	3.209	1.298e-01	1135	83.04	80.99	2.047	4.117e-01
conf.secret	1431	77.11	84.53	-7.417	6.506e-04	1133	76.57	84.67	-8.104	7.294e-04

Appendix D: Study replication (Genetic Matching)

Table 2 - Balance Statistics:

	ev	tv	diff	pvalue	ev	tv	diff	pvalue
age.group	2.5	2.4	0.032	0.52	2.5	2.4	0.084	0.332
educ	4.8	4.1	0.629	0.00	4.8	4.2	0.537	0.000
white.collar	30.3	27.6	2.668	0.29	30.3	26.1	4.133	0.115
not.full.time	27.7	33.5	-5.784	0.02	27.7	31.6	-3.865	0.143
male	49.7	49.1	0.557	0.87	49.7	48.1	1.505	0.628
tech	4.2	3.9	0.274	0.00	4.2	4.0	0.195	0.002
pol.info	1.5	1.3	0.164	0.00	1.5	1.3	0.159	0.000

Table 3 - Causal Effect of e-voting:

	N	prop.ev	prop.tv	diff	pvalue	N	prop.ev	prop.tv	diff	pvalue
eselect.cand	1388	83.84	53.42	30.428	1.237e-34	1298	83.84	54.89	28.958	1.321e-29
eval.voting	1460	46.33	21.30	25.035	1.833e-22	1364	46.33	20.59	25.739	3.091e-21
easy.voting	1469	33.64	11.53	22.111	5.420e-22	1373	33.64	10.76	22.879	6.374e-21
agree.evoting	1409	84.14	62.44	21.705	2.864e-20	1314	84.14	63.11	21.026	8.196e-18
how.clean	1284	57.97	40.99	16.980	2.561e-09	1196	57.97	38.82	19.157	1.809e-10
sure.counted	1418	86.35	77.02	9.332	7.444e-06	1327	86.35	77.03	9.315	1.895e-05
capable.auth	1416	85.14	76.25	8.889	2.954e-05	1321	85.14	76.33	8.804	7.624e-05
speed	1443	84.06	80.85	3.209	1.298e-01	1352	84.06	79.16	4.898	2.776e-02
conf.secret	1431	77.11	84.53	-7.417	6.506e-04	1337	77.11	84.42	-7.310	1.554e-03

Appendix E: Replication and Extension code in R

The link to the code, and the dataset can be found here: http://bit.ly/cs112-final-andriy-marta.

Appendix F: Estimation of LATE (Local Average Treatment Effect)

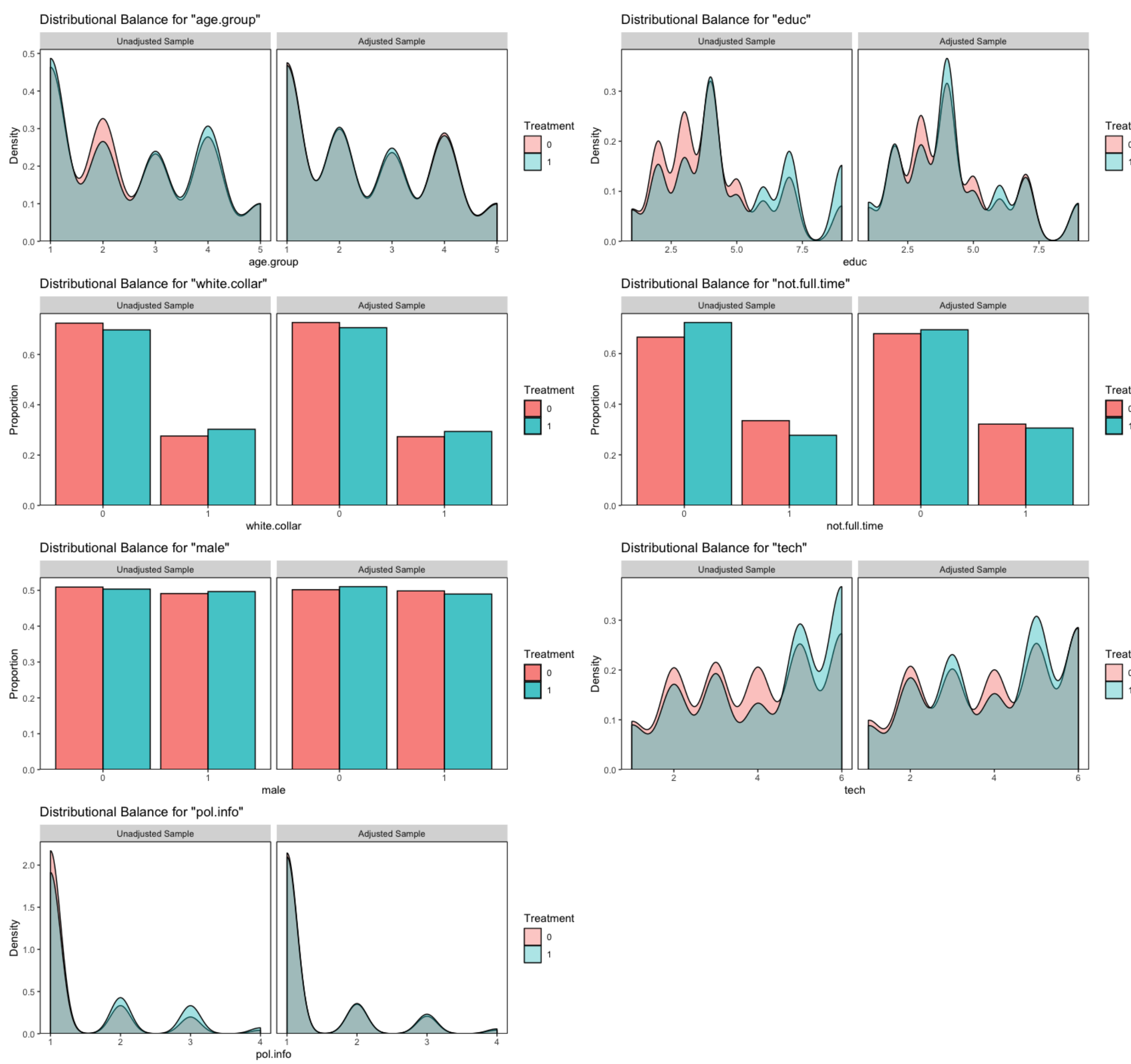
Formula to calculate the Local Average Treatment Effect:

$$LATE = ITT / CR$$

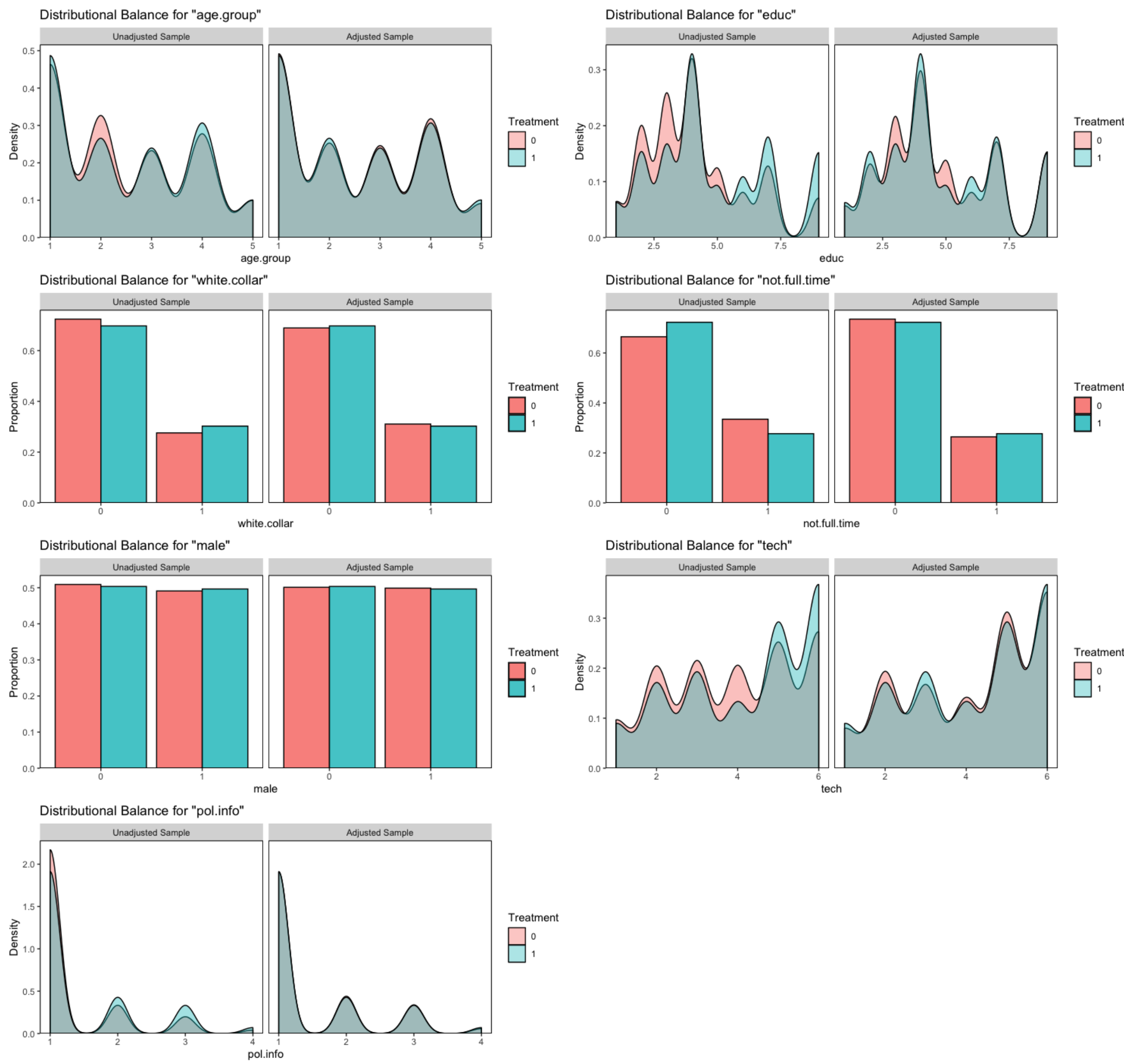
- *LATE* is computed by dividing the *Intention-To-Treat* effect by *Compliance Rate*.
- *ITT* (the effect of the encouragement on outcomes) can be calculated as the outcome of the encouraged group - (minus) the outcome of the non-encouraged group.

- *Compliance Rate* = Fraction of Subjects that were treated in the encouraged group - Fraction of Subjects that were treated in the non-encouraged group.
- Note: if the Compliance Rate is 100% (everyone complies with their assignment), the *Local Average Treatment Effect (LATE) = Intention-To-Treat effect*.
- The assumptions to hold true for the LATE to give an unbiased estimate are outlined in the section on "Suggestion for the future study."

Appendix G: Covariates balance plots before and after propensity score matching



Appendix H: Covariates balance plots before and after genetic matching



References

Alvarez, R. M., Levin, I., Pomares, J., & Leiras, M. (2013). Voting made safe and easy: the impact of e-voting on citizen perceptions. *Political Science Research and Methods*, *1*(1), 117–137. https://doi.org/10.1017/psrm.2013.2.

Diamond, A., Hainmueller, J. (2007). The Encouragement Design for Program Evaluation. IFC (Unpublished). Retrieved from https://drive.google.com/open?id=0B7xb7ljgS9qYT2hrcHpST0x4NEQ4Y0tDb3lqQlpZOUFIVlp3.

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics*, *95*(3), 932–945.

Election Guide. *Elections: Argentina Senate 2021*. IFES Election Guide. https://www.electionguide.org/elections/id/3593/.

Flannery, N. P. (2020, December 14). *Political Risk Analysis: How Serious Is Argentina's Economic Crisis?* Retrieved from https://www.forbes.com/sites/nathanielparishflannery/2020/12/15/political-risk-analysis-how-serious-is-argentinas-economic-crisis/?sh=3b143d038399.

Hadavand, A. (2020). The LATE estimator. Retrieved from https://www.notion.so/The-LATE-estimator-a2ad7662247547ffbbbec2eb715c4f50